AD-A048 218   YALE UNIV NEW HAVEN CONN DEPT OF STATISTICS   F/G 12/1
                HARMONIC REGRESSION.(U)
   DEC 77   F J ANSCOMBE   N00014-75-C-0563
UNCLASSIFIED   TR-44   NL

1 OF 1

ADA048 218

END
DATE
FILMED
1 -78
DDC

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD A048218

YALE UNIVERSITY

Department of Statistics

Technical Report, No. 44

December, 1977

26p.

Harmonic regression

by Francis J. Anscombe

TR-44

402 200

AD No.
DDC FILE COPY

DDC
RECEIVED
DEC 28 1977
F

HARMONIC REGRESSION*

F. J. Anscombe, Yale University

[Invited address, October 26, 1977, to the Third ERDA Statistical Symposium,
at Battelle Pacific Northwest Laboratories, Richland, WA]

The calculations of ordinary regression analysis -- linear regression by the method of least squares -- have been correctly doable for a century and a half. There have been changes in computational methods used. There is plenty to discuss about regression -- is it appropriate for the data, what do the results mean? No doubt the calculations are sometimes of little value, but sometimes they are appropriate and lead to new understanding.

Regression analysis of time series has a much shorter history. There is a good deal of literature about it, but the literature often has the air of arm-chair meditation by a non-participant. My concern has been to implement principles that are in the literature, and devise a working procedure. Various practical difficulties have been encountered that do not seem to be discussed in the literature.

Does anyone need to do regression analysis of time series? Conflicting opinions are heard. Huge amounts of time-series material are being collected and stored relating to the environment (weather, pollution), the observations being made daily or even more frequently. Many economic series are developed for monthly or weekly or daily activities. I have worked with annual series, which are probably the least satisfactory material for this kind of study.

Some broad generalities are presented below, and an example is shown. The details are vital, but as they have been fully described elsewhere they are not given here. (See the references; the detailed presentation is referred to as Chap. 10.)

Formulation

We consider regression of one "dependent" variable on just one "independent" or predictor variable. (Methods extend, of course, but not without some difficulties, to several predictor variables.) All means will be supposed zero. Then ordinary linear regression can be formulated thus. We are given observations on pairs of variables, (x_i, y_i) for $i = 1, 2, \ldots, m$. We suppose that for all i

$$y_i = \beta x_i + \varepsilon_i , \tag{1}$$

where the errors $\{\varepsilon_i\}$ are considered to be (in some sense) <u>independent</u> of each other and of the predictor variable $\{x_i\}$. The method of least squares can be equated to the method of maximum likelihood when we suppose that the $\{\varepsilon_i\}$ are independent random variables identically distributed $N(0, \sigma^2)$.

How should regression of time series be formulated? We are given series $\{x_t\}$, $\{y_t\}$, where $t = 1, 2, \ldots, n$. We shall not suppose these series, nor the error series $\{\varepsilon_t\}$ when we introduce it, to consist of independent elements. We shall instead suppose the series to be <u>stationary</u>, that is, realizations of some kind of stationary stochastic process. (In practice the appearance of stationarity with zero mean is encouraged by subtracting a linear or other trend, usually after taking logarithms.) To correspond to (1), one might suggest

$$y_t = \beta x_t + \varepsilon_t .$$

But if the series are related, the relation may be not simultaneous.
One might have

$$y_t = \beta x_{t-j} + \varepsilon_t ,$$

for some integer lag j . But then one might as well postulate

$$y_t = \Sigma_j \, \beta_j \, x_{t-j} + \varepsilon_t , \qquad (2)$$

where j runs over some suitable set of integer values. This seems to

be the appropriate formulation for stationary processes, to correspond

to (1) for independent processes. The first member of the right side

of (2) represents a linear filtering of $\{x_t\}$.

There are two main approaches to trying to estimate the parameters

$\{\beta_j\}$ of the filter in (2).

Time-domain methods

One can try direct multiple regression of $\{y_t\}$ on $\{x_t\}$ and on

lagged versions of it, $\{x_{t-j}\}$ for various j . There is a difficulty

about deciding how many lags should be considered. If $\{x_t\}$ is strongly

autocorrelated, conditioning will be poor. An accurate representation

of the relation between two stationary stochastic processes could easily

involve a large number of nonzero coefficients $\{\beta_j\}$.

If our reason for trying to fit a relation like (2) is to be able

to forecast y_t from past values of $\{x_t\}$, possibly a very crude esti-

mate of the $\{\beta_j\}$ will be good enough. The precision of a forecast is

limited by the variance of the error term. The greater precision that

would be attained if the $\{\beta_j\}$ were known exactly may be only negligibly

greater (Cleveland, 1967). Box and Jenkins (1970, 1976) have presented

a set of practical procedures for estimating the structures of time series well enough for forecasting. If our purpose is not forecasting, but understanding as well as we can the relation between the series, the Box-Jenkins methods may be less satisfactory.

It will be argued that some of these difficulties are mitigated or avoided by frequency-domain methods. However, we must usually be alert to temporal instability or change in a relation like (2), and that will be detected by time-domain methods.

Frequency-domain methods

The idea is to Fourier-transform the relation (2), and estimate the transform of $\{\beta_j\}$. It will be suggested that (i) this procedure is easier to carry out than multiple regression in the time domain, and that (ii) the results are easier to understand. Claim (i) derives from the fact that the first member on the right side of (2), the filtering of $\{x_t\}$, is a convolution of $\{\beta_j\}$ and $\{x_t\}$, and transforms to the product of the separate transforms of $\{\beta_j\}$ and $\{x_t\}$. Thus (2) becomes

$$FT\{y_t\} = (FT\{\beta_j\}) \cdot (FT\{x_t\}) + FT\{\epsilon_t\} .$$

These Fourier transforms are complex-valued functions of a real variable λ representing frequency. Consider a narrow frequency band (interval for λ). Suppose that in this interval the transform of $\{\beta_j\}$ were (near enough) constant. Then in this interval the relation between $FT\{y_t\}$ and $FT\{x_t\}$ would be exactly like relation (1) above between $\{y_i\}$ and $\{x_i\}$, with the exception that the variables and the regression coefficient are complex-valued. In real terms, $FT\{\beta_j\}$ is con-

veniently expressed as an amplitude, the gain function $G(\lambda)$, and an angle, the phase-shift function $\phi(\lambda)$. Thus if $G(\lambda)$ and $\phi(\lambda)$ could be regarded as constant over the frequency band, they could be estimated from the transforms of $\{y_t\}$ and $\{x_t\}$ by a slight modification of the usual procedure for the linear regression relation (1) -- expressed in real terms it looks a bit different, but the procedure is really ordinary linear least squares with two real coefficients to be estimated. The least-squares procedure is particularly appropriate if the error process $\{\varepsilon_t\}$ is a stationary Gaussian process whose spectral density is nearly constant over the band.

However, it has been generally recognized (Akaike and Yamanouchi, 1962, Cleveland and Parzen, 1975, and other writers) that treating $\phi(\lambda)$ as constant is not satisfactory when its derivative is much different from 0, and that it is better to approximate the behavior of the transform of $\{\beta_j\}$ in the narrow frequency band by three real parameters, the average values of $G(\lambda)$, $\phi(\lambda)$ and $\phi'(\lambda)$ in the band -- that is, treat $G(\lambda)$ as constant and $\phi(\lambda)$ as linear in λ . Now the regression procedure is further modified, becoming in fact nonlinear and requiring an iterative solution, but still computationally rather easy.

Thus the complete procedure involves examining the frequency range of λ in bands, using a moving "window", and in each band doing a small computation to determine three real parameters, representing average values of $G(\lambda)$, $\phi(\lambda)$, $\phi'(\lambda)$. Upon putting the solutions together we see the whole behavior of $G(\lambda)$ and $\phi(\lambda)$. With $G(\lambda)$ and $\phi(\lambda)$ estimated, $\{\beta_j\}$ could be inferred by making the inverse Fourier transform.

Intelligibility

Claim (ii) is that $G(\lambda)$ and $\phi(\lambda)$ are what we need, in order to understand the relation between $\{y_t\}$ and $\{x_t\}$, rather than the $\{\beta_j\}$. If the latter were given, we should have to Fourier-transform them to see qualitatively the effect of the filter. Compare with the usual commercial description of performance of an amplifier in a sound-reproduction system.

Example

As an example of methods, we try interrelating an annual series of total copper-mine output for the U.S. and two economic annual series, one giving the New York price of copper, the other the total dollar value of imports of merchandise into the U.S. The copper price series is thought to reflect the world supply and demand for copper. Changes in price might be expected to lead to similar changes in production, possibly a little later. The imports series is taken as an indicator of the U.S. economy. The copper production series is N235, and the price series is N241, in Historical Statistics of the United States (1975); the production figures run from 1845 to 1970, the prices from 1850 to 1970. The figures have been taken exactly as published, except that to smooth a change in price definition in 1968 the average of two definitions has been used for 1967. The price figures for 1850-1859 are of uncertain meaning, and the production figures for before 1860 show a more rapid proportional rate of growth than later. For present purposes it has seemed wise to ignore the pre-1860 data. Continuation of the series from 1970 to 1975 has been obtained from the Statistical Abstract of the United States (editions for 1975 and 1976). The two series are

reproduced in Figure 1, except that the last two digits of the production entries have been dropped for ease of reading. The imports series has been given in Chap. 10 and is not reproduced here; only the portion from 1860 to 1975 is used.

Figure 2 shows a plot against the date of the logarithm of the pro- duction series, with linear regression on date subtracted. Figure 3 is a similar plot for the price series. A plot for the imports series has been given in Chap. 10.

The three given series have each 116 entries (for 1860-1975). To prepare them for Fourier analysis they have been prewhitened by the steps: (i) take logarithms, (ii) subtract the linear regression on date, (iii) filter by the two-point filter with weights $(-0.9, 1)$. The last operation reduces the length of each series to 115. Then the series have been circularized (tapered) by linearly splicing the first 7 and the last 7 entries, so that the length of each series becomes 108. The Fourier transform is made at frequencies $(0, 1, 2, \ldots, 54)/108$ cycles per year; the transform is expressed as a set of (real) coefficients of cosine and sine terms, or alternatively as a set of squared amplitudes and phase angles. The frequencies are referred to as harmonics, num- bered 0 through 54.

The first step to perceiving an interrelation between any pair of series is to plot the difference of phase angles at each harmonic against the harmonic number. Figure 4 shows this for the production and price series, and Figure 5 for the production and imports series. At each harmonic, the product of the amplitudes is classified by size into one of six categories and represented by one of the plotting symbols:

.   ∘   O   ⊖   ▢   ⊞

Each phase difference is plotted with this symbol twice over in the
interval from 0 to 8 rightangles. In looking for trends, the viewer's
eye should be guided by the heavier symbols.

Figure 5 shows a fairly strong relation between production and
imports, especially at the higher frequencies -- the phase differences
are mostly rather close to 4 (or 0 or 8) rightangles, and show no trend
with frequency. A simultaneous positive correlation between these two
series is indicated. Figure 4 shows a less clear relation between pro-
duction and price. At lower frequencies there is some suggestion of
trend in the phase differences, implying that production follows price,
possibly by two years, possibly by four. At higher frequencies the
phase differences seem very scattered. Not reproduced is a phase dif-
ference plot for the price and imports series, suggesting quite a strong
simultaneous correlation at the lower frequencies, and not much at higher
frequencies.

Now the regression calculation in frequency bands, to estimate
$G(\lambda)$, $\phi(\lambda)$ and $\phi'(\lambda)$, can be performed. The window chosen is 23
harmonics wide, and sine weights have been used. The results are tabu-
lated in Figure 6. The first column lists the harmonic number of the
central frequency in the band; we have stepped the central harmonic
number from the lowest possible value, 11, by unit steps to the greatest
possible value, 43. (Had there been many more harmonics and a greater
bandwidth, greater steps would have been convenient.) The next three
columns list estimates of spectral density for, respectively, copper pro-
duction, copper price and imports (prewhitened as explained above),
obtained from the raw line spectra by the 23-point sine-weighted moving
average. The next four columns refer to regression of copper production

on copper price. They list average values in the band of $G(\lambda)$, $\phi(\lambda)$ and $\phi'(\lambda)$, and (in the fourth of these columns) multiple R^2 (the coherency). The behavior of $\phi(\lambda)$ and of R^2 gives a numerical measure of the trend seen in Figure 4. The last four columns of Figure 6 give similar information for regression of copper production on imports, and relate to Figure 5. (Simultaneous regression of copper production on both copper price and imports is not considered at this point.)

To test a null hypothesis of no association between series, 5%, 1% and 0.1% values for R^2 for any given frequency band are estimated (by a crude argument) at 0.19, 0.27 and 0.36, respectively -- these values probably err in being a little too low. The tabulated values are very highly correlated, as one reads down the column. So for regression of production on price, it seems reasonable to claim a substantial correlation at low frequencies, in the bands centered between the 11th and 19th harmonics. For regression of production on imports, the correlation is substantial in bands centered between the 23rd and 43rd harmonics -- R^2 is close to 0.5 in many of these bands.

Of our two predictor variables, copper price and general imports, the latter has on the whole the greater correlation with copper production. But the two predictor series have some correlation with each other. How useful is the price series as a predictor in conjunction with the imports series? Residual Fourier transforms of the production series and of the price series, after regression on the imports series, can be obtained, and a phase-difference plot can be made, analogous to Figure 4 for the original Fourier transforms. This is shown in Figure 7. The phase trend seems rather similar to that in Figure 4 at lower frequencies, and weaker at higher frequencies.

Figure 8 shows a calculation like that in Figure 6, but relating
to simultaneous regression of production on both price and imports,
instead of to separate regressions. The R^2 in the final column is
always greater than either value of R^2 (for the same frequency band)
given in Figure 6. The most striking increase over the R^2 for regres-
sion on imports only occurs for bands centered between the 25th and 28th
harmonics -- 0.479 instead of 0.332 at the 25th harmonic, 0.511 instead
of 0.359 at the 26th harmonic, etc. The same sort of crude argument as
before indicates that these four increases (but none of the others) can
be regarded as significant at the 5% level. The increases, on the
whole, are larger at lower frequencies than at higher frequencies.

The two phase-shift functions estimated in Figure 8 can be fairly
well approximated at most frequencies by saying that production is
correlated positively with imports of the same year and negatively with
prices of four years before.

Figures 9 and 10 are time-domain plots intended to show whether the
relations between the series perceived in the harmonic analysis pervade
the whole series or are special to particular epochs. For both plots,
the original series have been transformed to logarithms and a linear
trend has been subtracted. Then for Figure 9, low frequencies have
been suppressed by taking the second difference of the series, and the
resulting production values are plotted against the imports values. The
correlation coefficient is 0.60. The decade of each plotted point is
shown by the letters appearing on the right side of Figure 1; a star
means that two or more points have coincided. For Figure 10, the spectra
have been roughly whitened by taking the first difference of each
series, and then high frequencies have been suppressed by three simple

2-point averagings. The first 4 values of the resulting production series and imports series have been dropped, and the last 4 values of the resulting price series; and then the production values are plotted against the linear combination of the imports values (for the same year) minus 0.8 times the price values (for four years earlier). The correlation coefficient is 0.53. The decade of the production values is shown as before.

The pronounced correlation in both Figures 9 and 10 is due to a few extreme points labeled G or H , representing the two decades from 1920 to 1939. If all points for these decades were omitted, the correlation would become 0.04 for Figure 9 and -0.08 for Figure 10. That is, the correlation would disappear.

Discussion

Harmonic regression is a systematic way of looking for association between series in all parts of the frequency range. It is unlikely to reveal anything that cannot be found by careful visual comparison of plots such as those in Figures 2 and 3, at least when only two or three series are under consideration. (A similar remark can be made about ordinary regression.)

We have found clear evidence of association between copper production and general imports, at higher frequencies, and some suggestion of predictive value for copper price also, at middle-to-low frequencies. What associations there are seem to inhere in the economically turbulent years of the 20's and 30's. We do not see similar associations in the other decades. Possibly relations between these series are changing, possibly the phenomena are highly nonlinear.

References

The detailed study on which this paper is based is

F. J. ANSCOMBE (in preparation). Statistical Computing with APL, Chap. 10: "Time series: Yale enrolment".

Other references are

H. AKAIKE and Y. YAMANOUCHI (1962). On the statistical estimation of frequency response function. Annals of the Institute of Statistical Mathematics, 14, 23-56.

G. E. P. BOX and G. M. JENKINS (1970, 1976). Time Series Analysis: Forecasting and Control. Holden-Day.

W. S. CLEVELAND (1967). Time series projection: theory and practice. Doctoral dissertation, Yale University.

W. S. CLEVELAND and E. PARZEN (1975). The estimation of coherence, frequency response, and envelope delay. Technometrics, 17, 167-172.

U. S. Bureau of the Census (1975). Historical Statistics of the United States, Colonial Times to 1970, Bicentennial Edition. U. S. Government Printing Office.

+	0	1	2	3	4	5	6	7	8	9	

U.S. COPPER PRODUCTION (MINE OUTPUT, HUNDREDS OF SHORT TONS)

	0	1	2	3	4	5	6	7	8	9	
1860	81	84	106	95	90	95	100	112	130	140	A
1870	141	146	140	174	196	202	213	235	241	258	B
1880	302	358	453	578	725	829	789	907	1132	1134	C
1890	1299	1421	1725	1647	1771	1903	2300	2470	2633	2843	D
1900	3031	3010	3298	3490	4063	4444	4585	4236	4784	5633	E
1910	5441	5574	6245	6178	5742	7440	10029	9477	9550	6062	F
1920	6123	2331	4823	7389	8031	8391	8626	8250	9049	9976	G
1930	7051	5289	2381	1906	2374	3865	6145	8420	5578	7283	H
1940	8781	9581	10801	10908	9725	7729	6087	8476	8348	7528	I
1950	9093	9283	9254	9264	8355	9986	11042	10869	9793	8248	J
1960	10802	11652	12284	12132	12468	13517	14292	9541	12046	15446	K
1970	17197	15220	16650	17180	15970	14110					L

PRICE OF REFINED COPPER AT NEW YORK (CENTS PER POUND)

	0	1	2	3	4	5	6	7	8	9	
1860	22.88	22.25	21.88	33.88	47.00	39.25	34.25	25.38	23.00	24.25	A
1870	21.19	24.12	35.56	28.00	22.00	22.69	21.00	19.00	16.56	18.62	B
1880	21.50	18.25	18.50	15.88	13.75	11.10	11.00	11.25	16.80	13.75	C
1890	15.75	12.88	11.50	10.65	9.43	10.70	10.92	11.30	12.01	17.75	D
1900	16.54	16.40	11.96	13.62	13.11	15.98	19.77	20.86	13.39	13.11	E
1910	12.88	12.55	16.48	15.52	13.31	17.47	28.46	29.19	29.19	18.90	F
1920	17.50	12.65	13.56	14.61	13.16	14.16	13.95	13.05	14.68	18.23	G
1930	13.11	8.24	5.67	7.15	8.53	8.76	9.58	13.27	10.10	11.07	H
1940	11.40	11.87	11.87	11.87	11.87	11.87	13.92	21.15	22.20	19.36	I
1950	21.46	24.37	24.37	28.92	29.82	37.39	41.88	29.99	26.13	30.82	J
1960	32.16	30.14	30.82	30.82	32.17	35.19	35.82	38.01	41.17	47.43	K
1970	58.07	52.00	51.20	59.50	77.30	64.20					L

FIGURE 1

COMMON LOGARITHM OF COPPER PRODUCTION--RESIDUALS FROM FITTED LINE

FIGURE 2

-16-



COMMON LOGARITHM OF COPPER PRICE--RESIDUALS FROM FITTED LINE

FIGURE 3

-18-
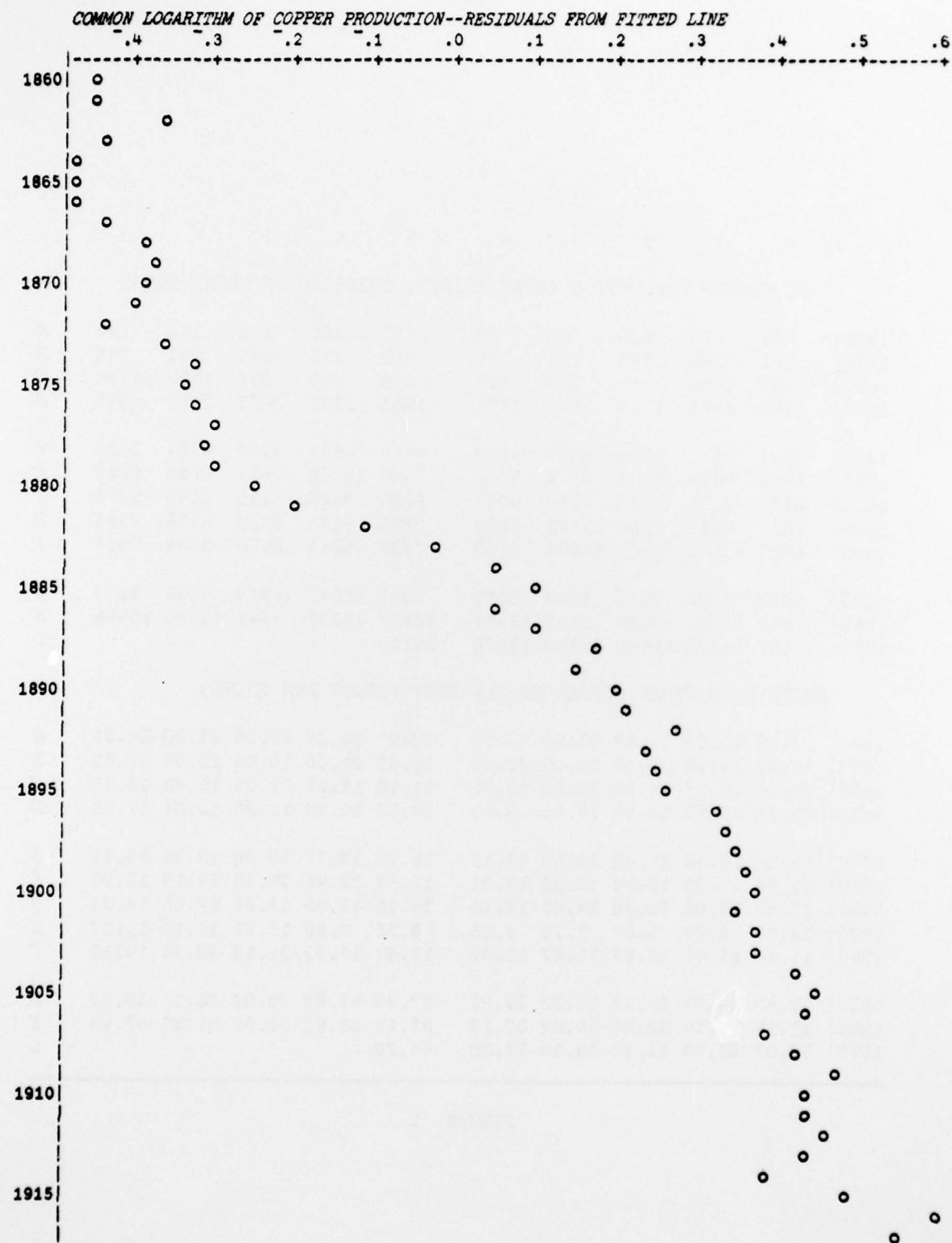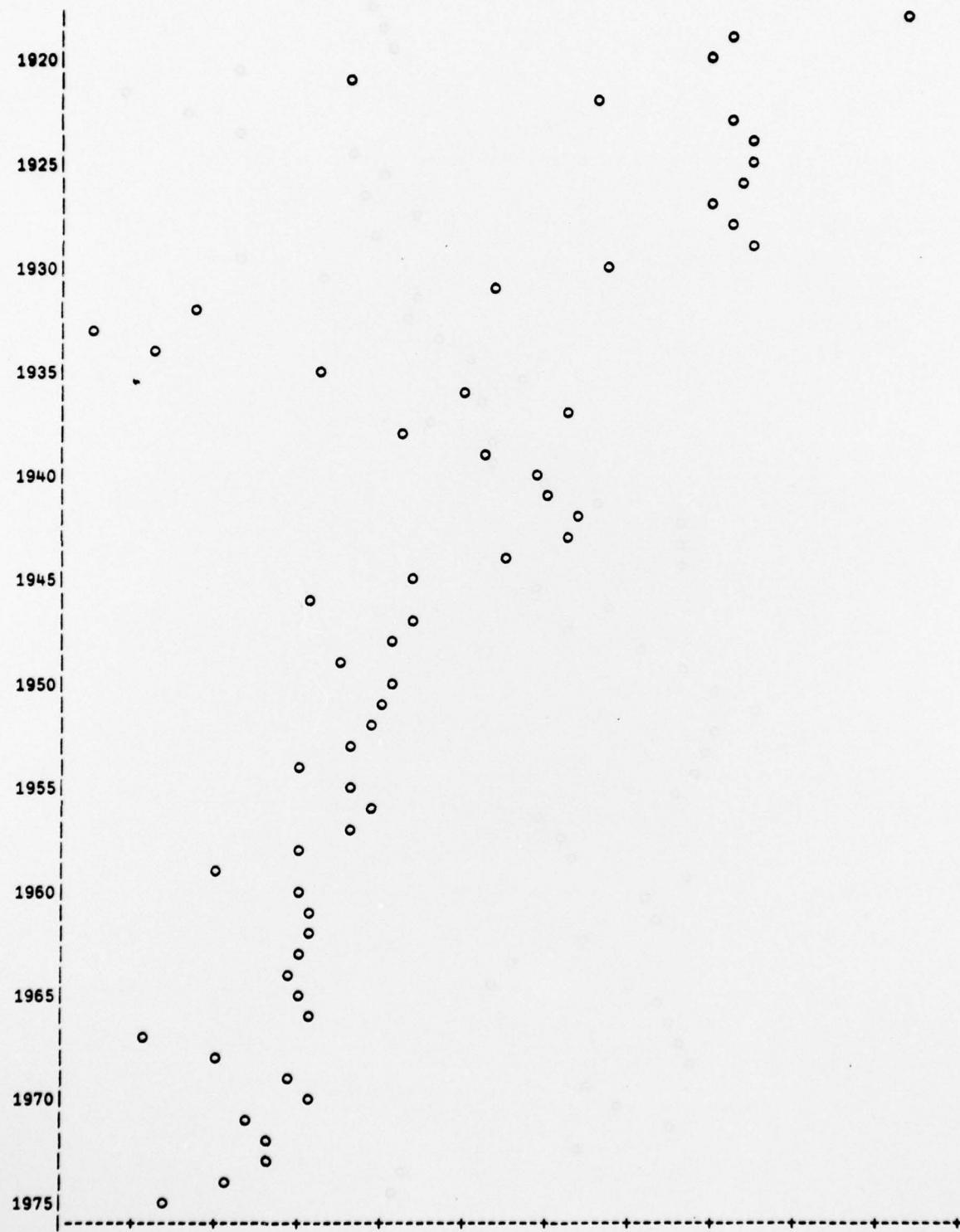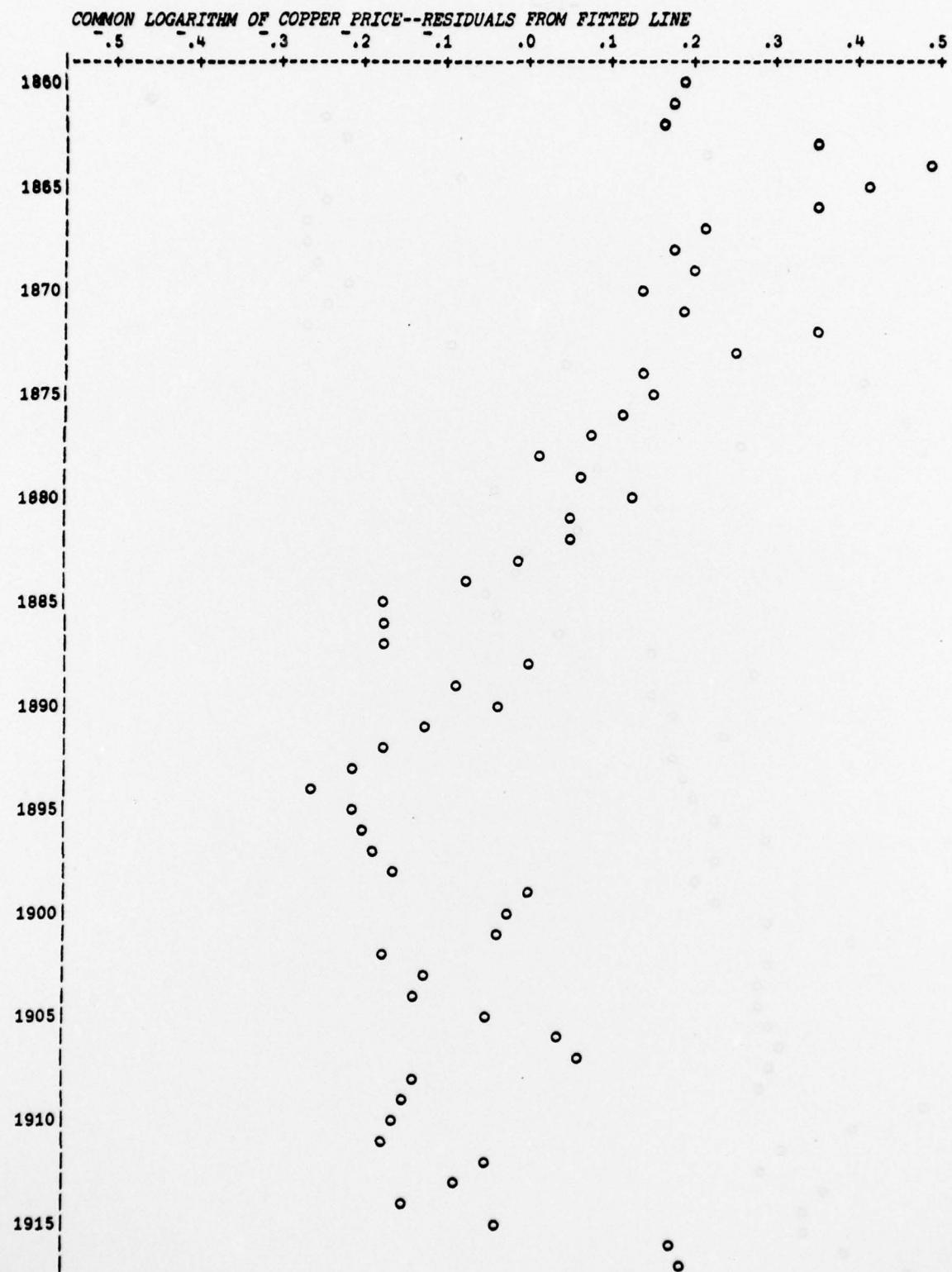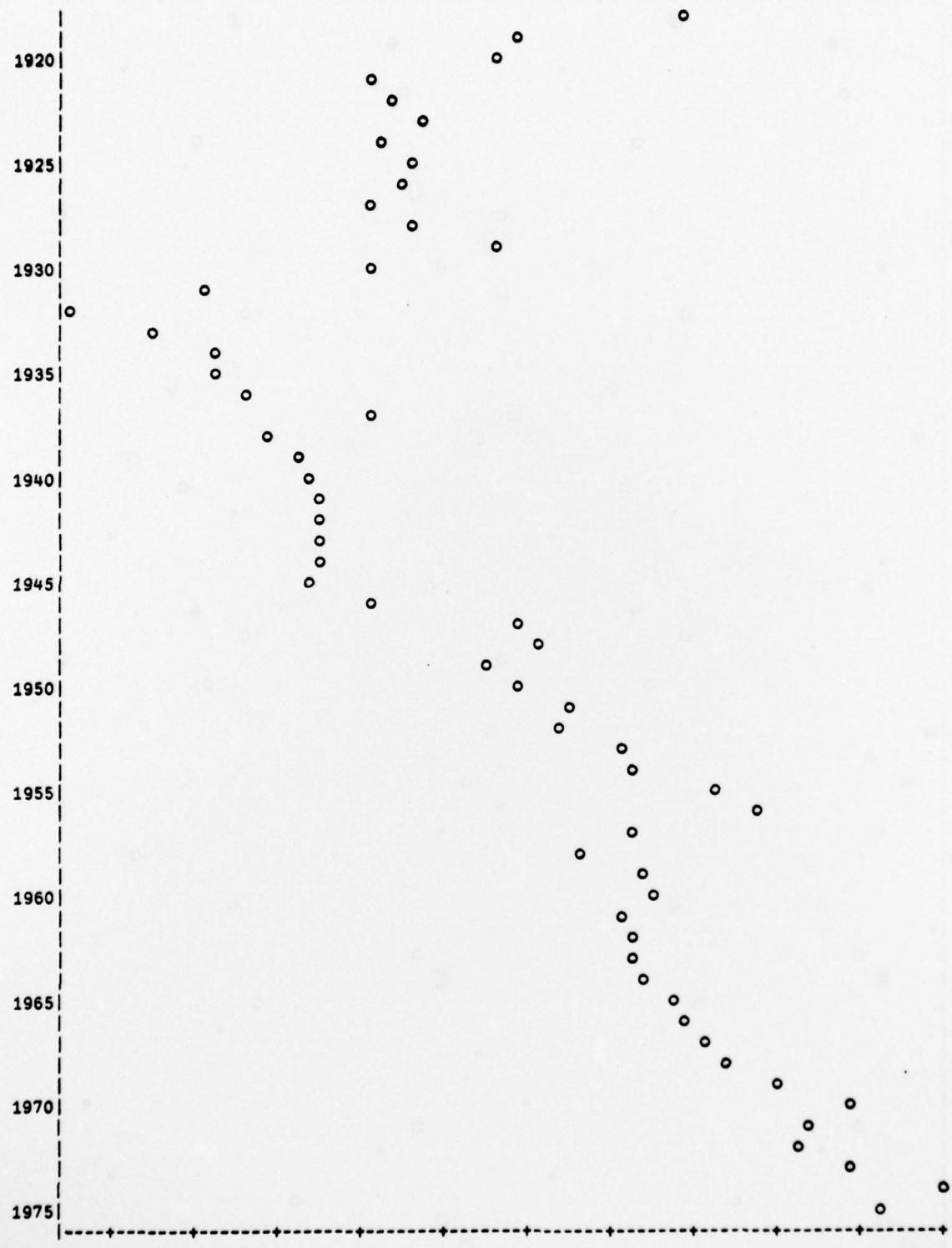
PHASE DIFFERENCE (RIGHTANGLES): COPPER PRODUCTION AND COPPER PRICE

FIGURE 4

PHASE DIFFERENCE (RIGHTANGLES): COPPER PRODUCTION AND GENERAL IMPORTS
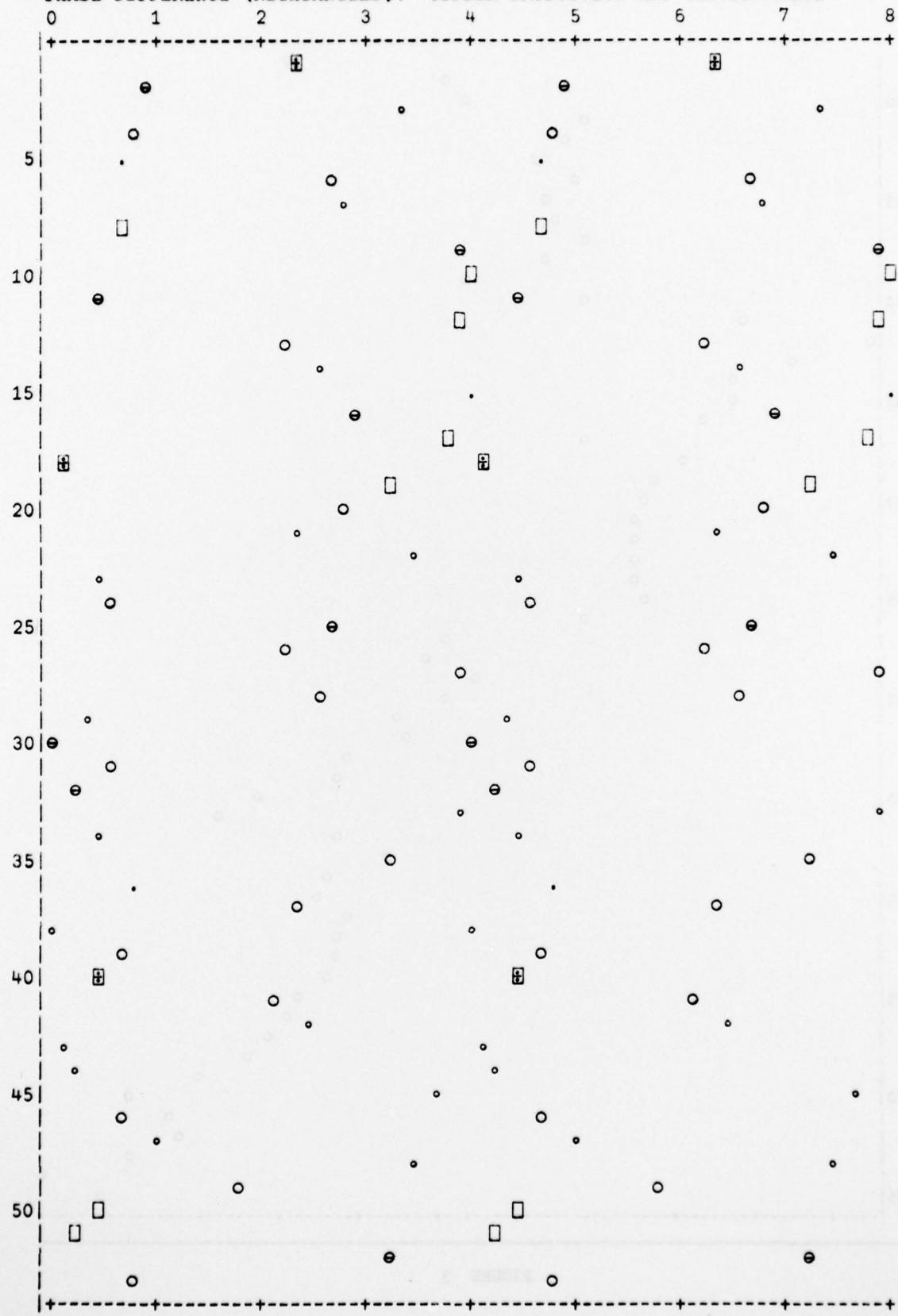
FIGURE 5
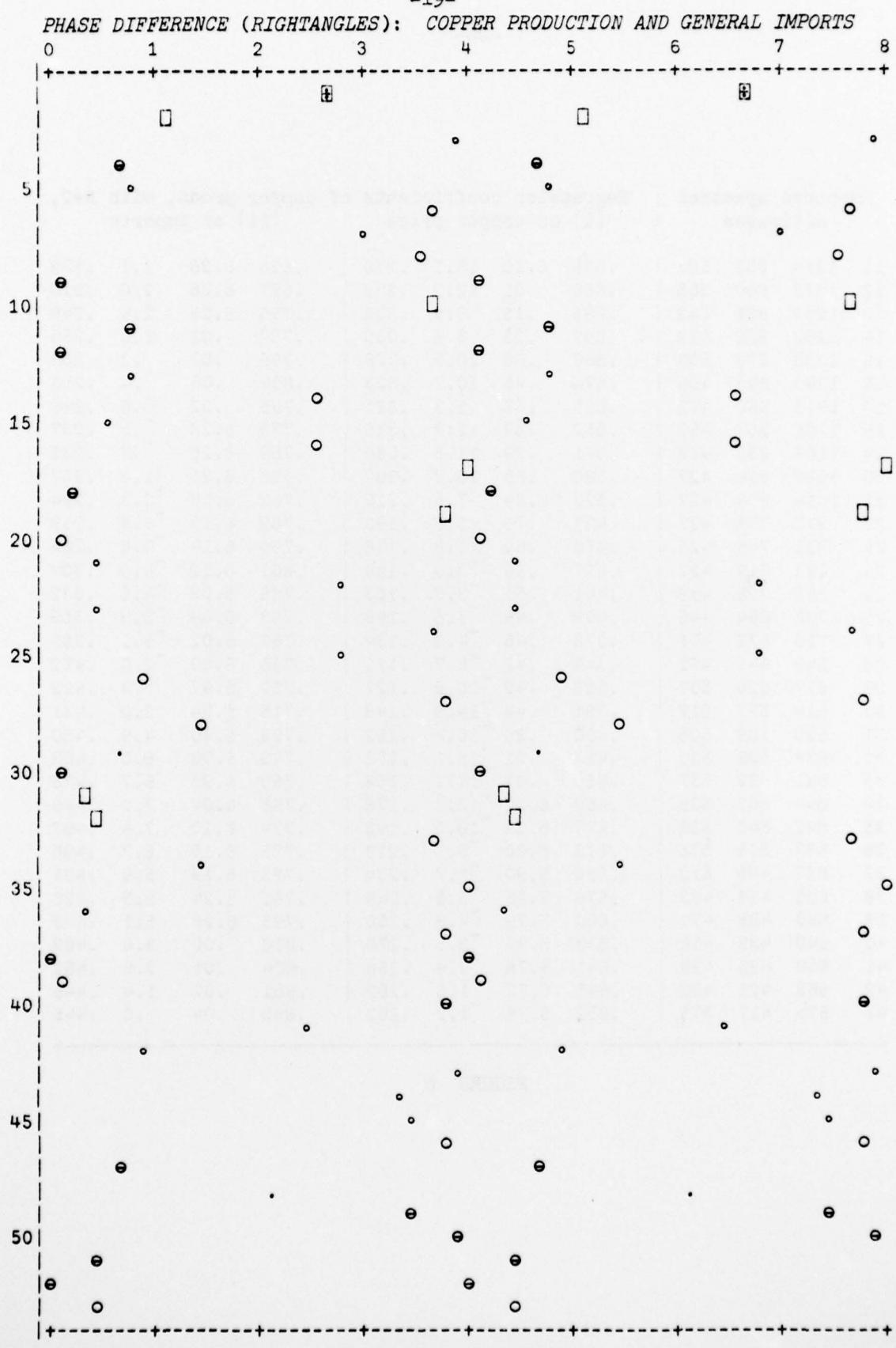
Smoothed spectral estimates			Regression coefficients of copper prodn, with $R*2$,							
			(i) on copper price				(ii) on imports			
11	1274	853	583	.635	6.10	15.1	.270	.625	6.26	1.1	.179
12	1272	860	566	.666	.01	12.2	.299	.687	6.26	2.0	.210
13	1252	861	542	.695	.15	9.8	.332	.756	6.26	2.4	.248
14	1259	868	519	.692	.25	9.6	.330	.787	.02	1.0	.255
15	1258	879	503	.686	.36	10.6	.328	.796	.03	.3	.253
16	1240	883	486	.674	.46	10.7	.323	.804	.03	-.4	.253
17	1213	880	472	.665	.57	11.3	.321	.795	.02	-.6	.246
18	1166	867	455	.652	.69	11.9	.316	.779	6.28	-.5	.237
19	1109	852	439	.621	.79	11.5	.296	.768	6.26	-.8	.233
20	1058	824	427	.580	.86	10.5	.262	.766	6.21	-1.8	.237
21	1014	805	427	.525	.84	7.5	.219	.762	6.18	-3.3	.244
22	972	777	427	.487	.70	1.8	.190	.769	6.15	-3.8	.259
23	936	745	425	.473	.60	-1.8	.178	.790	6.14	-4.8	.284
24	893	719	427	.457	.56	-3.0	.168	.801	6.12	-4.9	.307
25	842	706	439	.441	.51	-3.3	.163	.798	6.08	-4.6	.332
26	782	694	446	.409	.48	-3.6	.149	.793	6.04	-3.9	.359
27	716	672	471	.378	.46	-4.2	.134	.761	6.02	-3.1	.381
28	646	641	491	.349	.42	-5.7	.121	.736	5.99	-2.0	.412
29	617	620	507	.355	.42	-10.5	.127	.717	5.97	-.3	.422
30	614	597	517	.391	.44	-14.9	.149	.715	5.94	2.0	.431
31	620	589	525	.400	.36	-16.4	.152	.729	5.90	4.9	.450
32	634	589	531	.413	.21	-15.1	.158	.743	5.93	6.0	.463
33	641	582	537	.430	.07	-13.7	.168	.753	5.98	6.7	.475
34	644	567	536	.450	6.22	-12.3	.178	.766	6.04	7.2	.488
35	642	545	533	.477	6.11	-10.5	.193	.774	6.10	7.4	.497
36	637	515	526	.513	6.00	-8.7	.213	.775	6.15	6.7	.496
37	627	486	512	.550	5.90	-5.7	.234	.775	6.19	5.9	.491
38	625	458	492	.578	5.83	-5.5	.245	.781	6.24	5.5	.480
39	630	438	471	.600	5.79	-4.8	.250	.795	6.28	5.1	.473
40	640	432	450	.633	5.77	-3.5	.270	.810	.00	3.8	.462
41	659	425	428	.641	5.78	-2.4	.266	.834	.01	2.6	.452
42	669	421	402	.645	5.77	-1.6	.262	.861	.02	1.4	.446
43	675	417	375	.652	5.76	-1.2	.262	.895	.04	.0	.445

FIGURE 6

PHASE DIFFERENCE (RIGHTANGLES): PRODUCTION RESIDUALS AND PRICE RESIDUALS
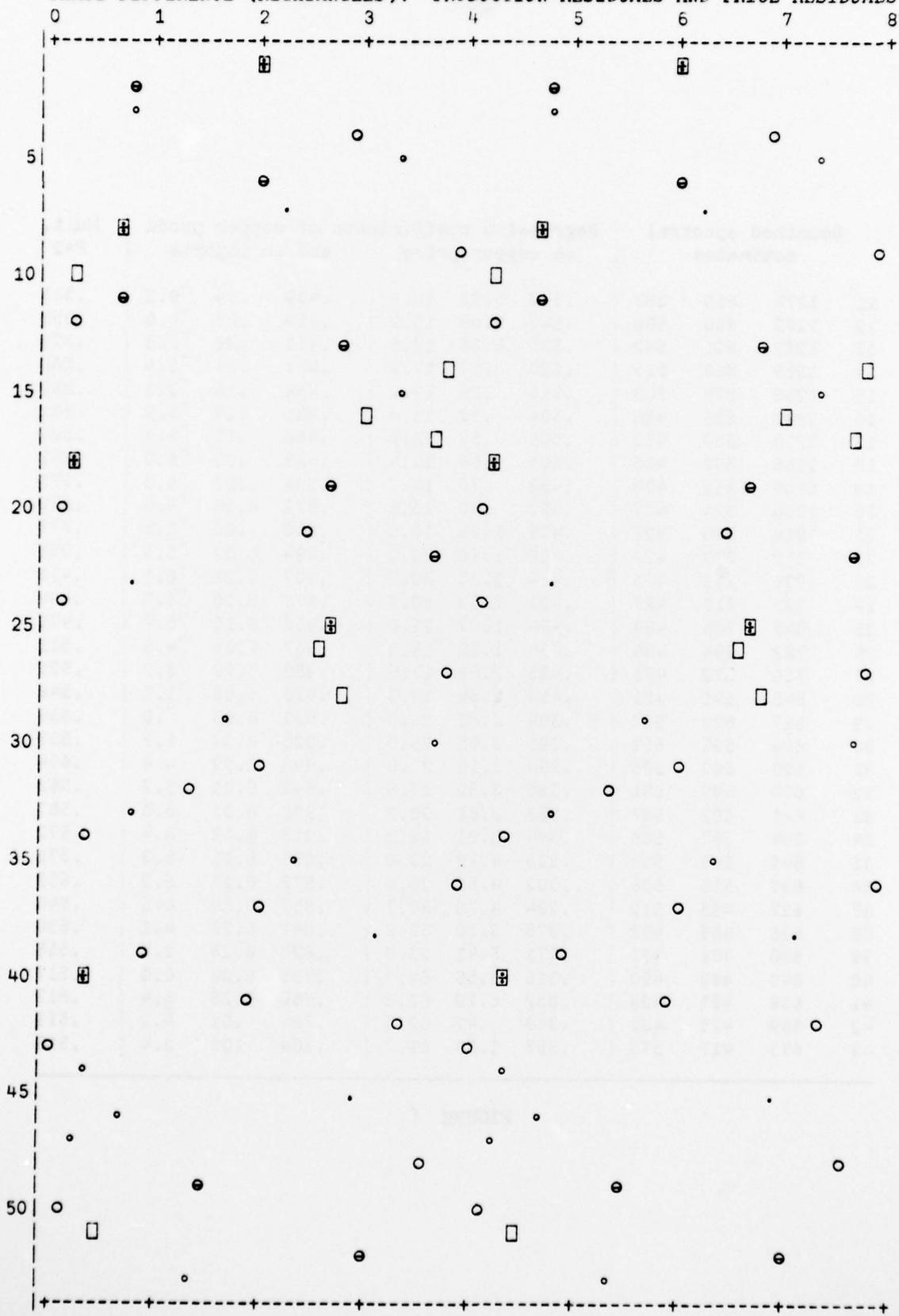
FIGURE 7

	Smoothed spectral estimates			Regression coefficients of copper prodn on copper price			and on imports			Mult. R*2
11	1274	853	583	.542	5.92	18.2	.439	.34	-6.2	.345
12	1272	860	566	.542	6.09	15.9	.424	.33	-4.6	.359
13	1252	861	542	.537	6.26	12.5	.411	.29	-1.3	.379
14	1259	868	519	.520	.12	12.2	.431	.24	-2.0	.380
15	1258	879	503	.515	.26	13.2	.439	.18	-3.1	.380
16	1240	883	486	.504	.39	13.4	.461	.15	-3.9	.381
17	1213	880	472	.506	.52	13.9	.466	.11	-4.4	.382
18	1166	867	455	.505	.66	14.4	.469	.05	-5.0	.382
19	1109	852	439	.482	.78	14.7	.504	6.28	-5.3	.377
20	1058	824	427	.450	.90	15.6	.572	6.26	-5.5	.372
21	1014	805	427	.429	1.01	18.0	.660	.00	-6.3	.378
22	972	777	427	.416	1.18	20.0	.734	6.27	-5.9	.391
23	936	745	425	.414	1.35	20.7	.807	6.25	-6.1	.415
24	893	719	427	.421	1.53	20.9	.871	6.20	-5.9	.444
25	842	706	439	.434	1.72	21.0	.919	6.15	-5.7	.479
26	782	694	446	.434	1.89	18.5	.957	6.04	-4.5	.511
27	716	672	471	.425	2.09	17.9	.952	5.99	-3.2	.525
28	646	641	491	.414	2.34	18.9	.950	5.98	-1.7	.548
29	617	620	507	.394	2.81	24.6	.923	6.06	-.0	.539
30	614	597	517	.391	3.05	25.0	.925	6.04	1.8	.537
31	620	589	525	.394	3.18	27.6	.943	5.99	4.4	.554
32	634	589	531	.385	3.39	27.8	.942	6.01	5.7	.563
33	641	582	537	.363	3.65	28.3	.925	6.05	6.0	.567
34	644	567	536	.344	3.91	28.8	.913	6.10	6.4	.572
35	642	545	533	.325	4.18	29.0	.896	6.15	6.3	.572
36	637	515	526	.302	4.51	29.5	.872	6.17	5.2	.559
37	627	486	512	.284	4.78	30.7	.857	6.19	4.5	.546
38	625	458	492	.275	5.10	30.9	.847	6.22	4.1	.530
39	630	438	471	.271	5.41	23.8	.823	6.28	2.8	.516
40	640	432	450	.315	5.56	64.2	.735	6.24	6.0	.517
41	659	425	428	.352	6.19	62.5	.766	6.26	5.4	.517
42	669	421	402	.359	.47	62.7	.784	.01	4.2	.513
43	675	417	375	.357	1.05	63.3	.804	.05	2.4	.513

FIGURE 8

Copper production vs. imports: high-frequency filtering.

FIGURE 9

```
                                                                    H

                                               *       G        H


                        C        C
                                     F    C
                   C       I   KK        C
          A          D   EB  H*I  DK  *  FCC B  G
                  BD BA E* A JJD* C B  CCIK B
0|          B      H      D J**EF E* I*KB G F E*  B
 |          H G      G J G K  FI *L *              B
 |            G         J         KH L F
 |                 *              I
 |          G                       K
 |                                 II
 |                                  F


 |              F
 |    H G
 |
 |       H
 |
|H
|--------------------------------------+--------------------------------------
                                        0
```

Copper production vs. imports and price: low-frequency filtering.
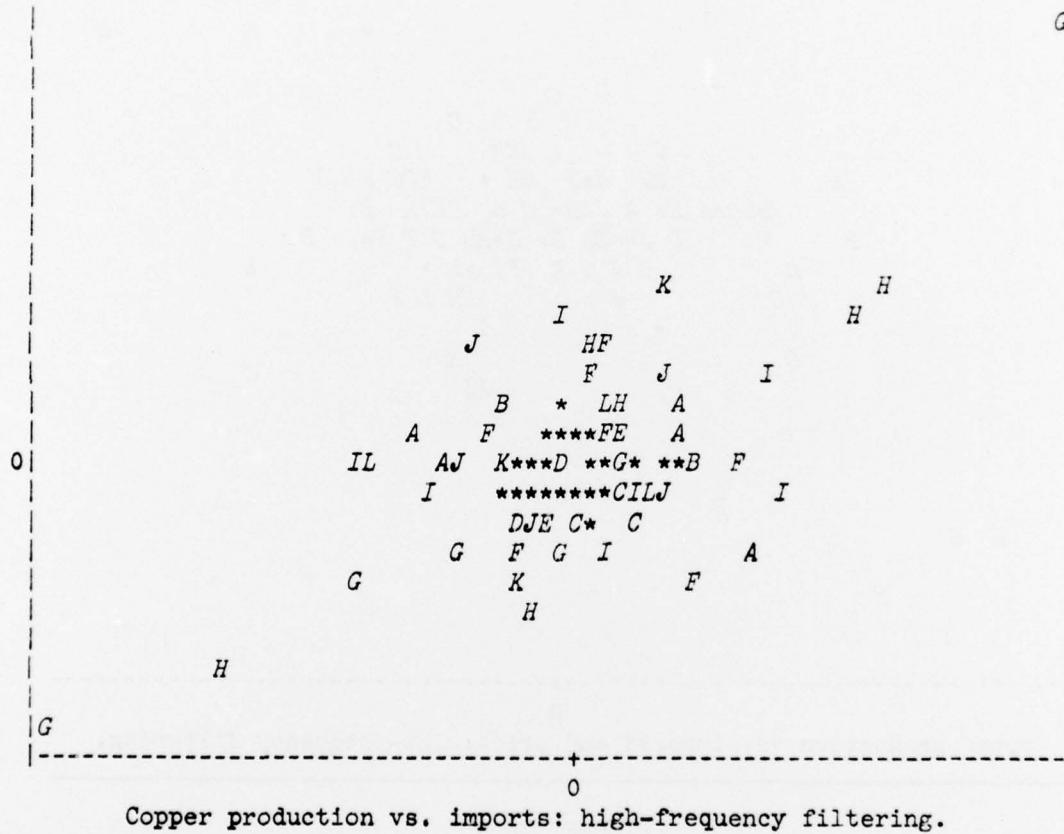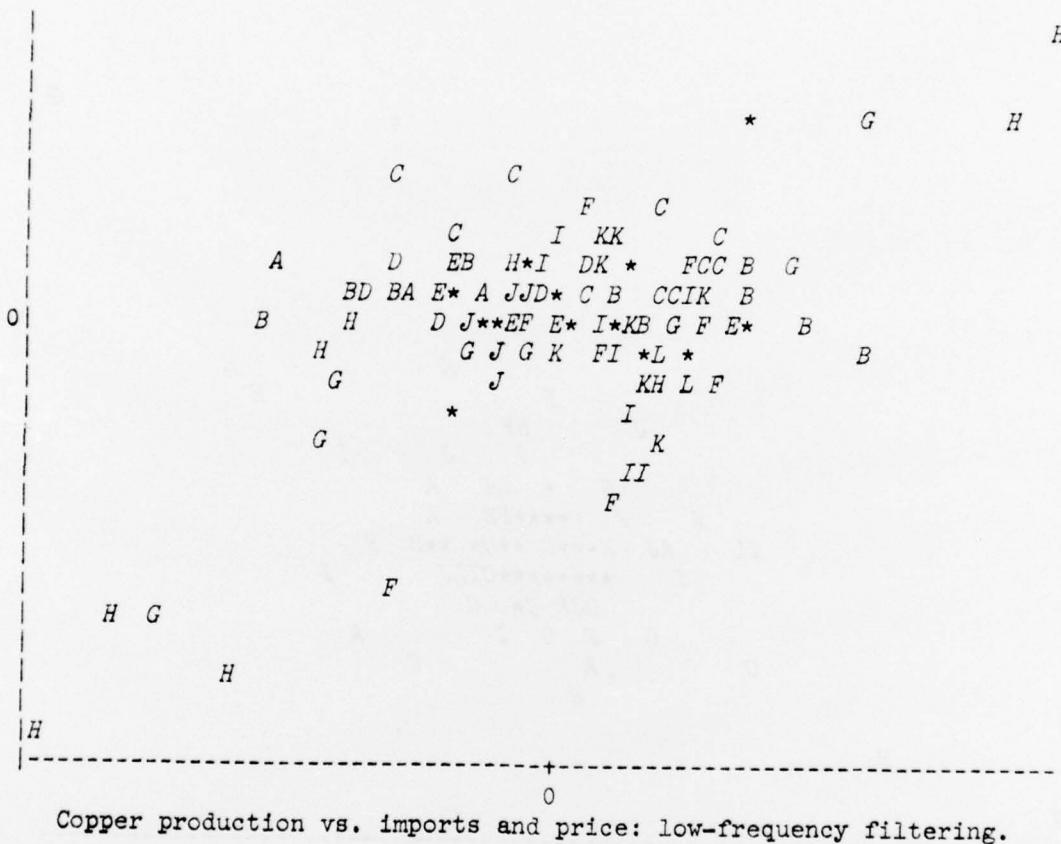
FIGURE 10

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 44	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) HARMONIC REGRESSION		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Francis J. Anscombe		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0563
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Yale University New Haven, CT 06520		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 042-242
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics and Probability Program (Code 436) Arlington, VA 22217		12. REPORT DATE December, 1977
		13. NUMBER OF PAGES i + 24 + ii = 27
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Time series analysis
Spectral analysis
Regression of time series
Copper production

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

See next page.

DD FORM 1 JAN 73 1473   EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601 |

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Abstract

Ordinary linear regression, by the method of least squares, is used to determine a linear relation subsisting between given independent observations of two or more variables. An analogous problem for time series is to determine a linear relation subsisting between two or more given stationary series. The linear relation may take the form that one series is a linear filtering of the other series, plus a stationary error process. The coefficients of the filter can be determined directly by multiple regression in the time domain, but there are difficulties. An easier procedure, leading to more intelligible results, is to estimate the Fourier transform of the coefficients of the filter for each predictor series (conveniently expressed as a gain function and a phase-shift function) by simpler regression calculations in the frequency domain. The procedure is illustrated by a study of interrelationship of an annual series of output of U.S. copper mines from 1860 to 1975, and two annual economic series relating to the same time period, namely a series of copper prices at New York and a series of total dollar value of general imports of merchandise into the U.S.